

MOTIVATIONS

- COVID-19: disruptions in daily life and work
- Twitter: public social platform for opinions
- More than 60 million Twitter users in the United States
- Sentiment: people's attitude toward COVID-19
- Potential topics:

Health, safety, economics, social distancing, and school reopenings

- Machine learning:

Latent Dirichlet Allocation (LDA): unsupervised learning for clusterings/topics

TWITTER DATA

- Twitter API: 8 days of data (Aug 08 to Aug 15)
- Limited to 5000 tweets per day related to COVID-19: English, within United States
- Search key words:


corona, #corona, coronavirus, #coronavirus, covid, #covid, covid19, #covid-19, sarscov2, #sarscov2, sars cov2, sars cov 2, covid_19, #covid_19, #ncov, ncov, #ncov2019, ncov2019, 2019-ncov, #2019-ncov, pandemic, #pandemic, #2019ncov, 2019ncov







- A total of 40K Tweets

Full tweet text and geo information




A SNAPSHOT FOR A TWEET



-  Home
-  Explore
-  Notifications
-  Messages
-  Bookmarks
-  Lists

← Tweet







Justin (FroYo'g Softserve 🍌 if you're Scoopy)
@FroyogSoftserve

Back in March, the day before I was due to get a filling, I busted a tooth on a cheezit. Extra toasty. Because of COVID, the dentist office was closed. Today, I went to get that filling, and now I'm getting a crown. FFS, FLOSS CHILDREN!


7:59 PM · Aug 14, 2020 from Brooklyn, OH · Twitter for Android

10 Likes



🔍 Search Twitter

Relevant people



Justin (FroYo'g Softs...
@FroyogSoftserve [Follow](#)


Eldritch being from non-Euclidian area of Ohio. Scoop. Wavelength on the autism spectrum. I dig atheism, fellow Scoops, and Shock Monkeys. Tweets/views=my own.

What's happening

US election · 1 hour ago

Trump campaign releases list of speakers for the RNC

Trending with: [The Honorable](#) and [Ivanka Trump](#)





A SNAPSHOT FOR A TWEET

```
{'created_at': 'Fri Aug 14 23:59:08 +0000 2020',
 'id': 1294423341001060352,
 'id_str': '1294423341001060352',
 'full_text': "Back in March, the day before I was due to get a filling, I busted a tooth on a cheezit. Extra toasty. Because of COV
ID, the dentist office was closed. Today, I went to get that filling, and now I'm getting a crown. FFS, FLOSS CHILDREN!",
 'truncated': False,
 'display_text_range': [0, 236],
 'entities': {'hashtags': [], 'symbols': [], 'user_mentions': [], 'urls': []},
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'source': '<a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>',
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'in_reply_to_screen_name': None,
 'user': {'id': 1708238947,
 'id_str': '1708238947',
 'name': "Justin (FroYo'g Softserve 🍌 if you're Scoopy)",
 'screen_name': 'FroyogSoftserve',
 'location': 'Brooklyn, OH',
 'description': 'Eldritch being from non-Euclidian area of Ohio. Scoop. Wavelength on the autism spectrum. I dig atheism, fellow Sc
oops, and Shock Monkeys. Tweets/views=my own.',
 'url': None,
 'entities': {'description': {'urls': []}},
 'protected': False,
 'followers_count': 970,
 'friends_count': 2163,
 'listed_count': 10,
 'created_at': 'Wed Aug 28 20:37:27 +0000 2013',
 'favourites_count': 45385,
 'utc_offset': None,
 'time_zone': None,
 'geo_enabled': True,
 'verified': False,
 'statuses_count': 10169,
 'lang': None,
```



A SNAPSHOT FOR A TWEET

```
'notifications': None,
'translator_type': 'none'},
'geo': None,
'coordinates': None,
'place': {'id': '91a74c998f238191',
'url': 'https://api.twitter.com/1.1/geo/id/91a74c998f238191.json',
'place_type': 'city',
'name': 'Brooklyn',
'full_name': 'Brooklyn, OH',
'country_code': 'US',
'country': 'United States',
'contained_within': [],
'bounding_box': {'type': 'Polygon',
'coordinates': [[[-81.7697, 41.418412],
[-81.728559, 41.418412],
[-81.728559, 41.462537],
[-81.7697, 41.462537]]]}},
'attributes': {}},
'contributors': None,
'is_quote_status': False,
'retweet_count': 0,
'favorite_count': 10,
'favorited': False,
'retweeted': False,
'lang': 'en'}
```

SENTIMENT ANALYSIS



SENTIMENT EXAMPLES

1	@greggutfeld Quit trying to be the mask police and you won't be assaulted or cussed out when I see someone with no mask I just turn and walk a different direction doing my social distancing Which I do anyway pandemic or not
-1	Sources: College Football "Is Done" Due To COVID-19 Concerns https://t.co/UNCgN58RjX via @TheSportsOnTap
-1	@GOP @realDonaldTrump Biden would have had COVID under control by now, had real medicine to help the sick and wouldn't be making phony promises about a vaccine.
1	@StCthrin For the first part of the Covid shutdowns, I made a different cocktail practically every day. Basically a pro now, thanks corona.
0	@CBSNews Screw Pelosi, it's always about Her , last Stimulus Democrats gave themselves \$38,000 Raise ➡ Which Had Nothing to Do with The Coronavirus to Help The American People!!😏



SENTIMENT ANALYSIS

- Sentiment analysis:

The process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining.

- TextBlob

A Python library for processing textual data, part of the natural language processing for part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

NATURAL LANGUAGE PROCESSING (NLP)

- NLP definition:

A subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.



NLP CONTINUED

- Splitted and stemmed (example):

*Before: "As Columbia County suffers a #COVID19 spike due to golfing events. ... a political golf event in Easton with NO MASKS and NO SOCIAL DISTANCING. #NewYork #MaskUp
<https://t.co/zeOe511KdM>"*

*After: "columbia counti suffer #covid19 spike due golf event ... polit golf event easton mask
social distanc #newyork #maskup http co zoe511kdm"*

- Final words count: >70K (18k words showed up more than once)

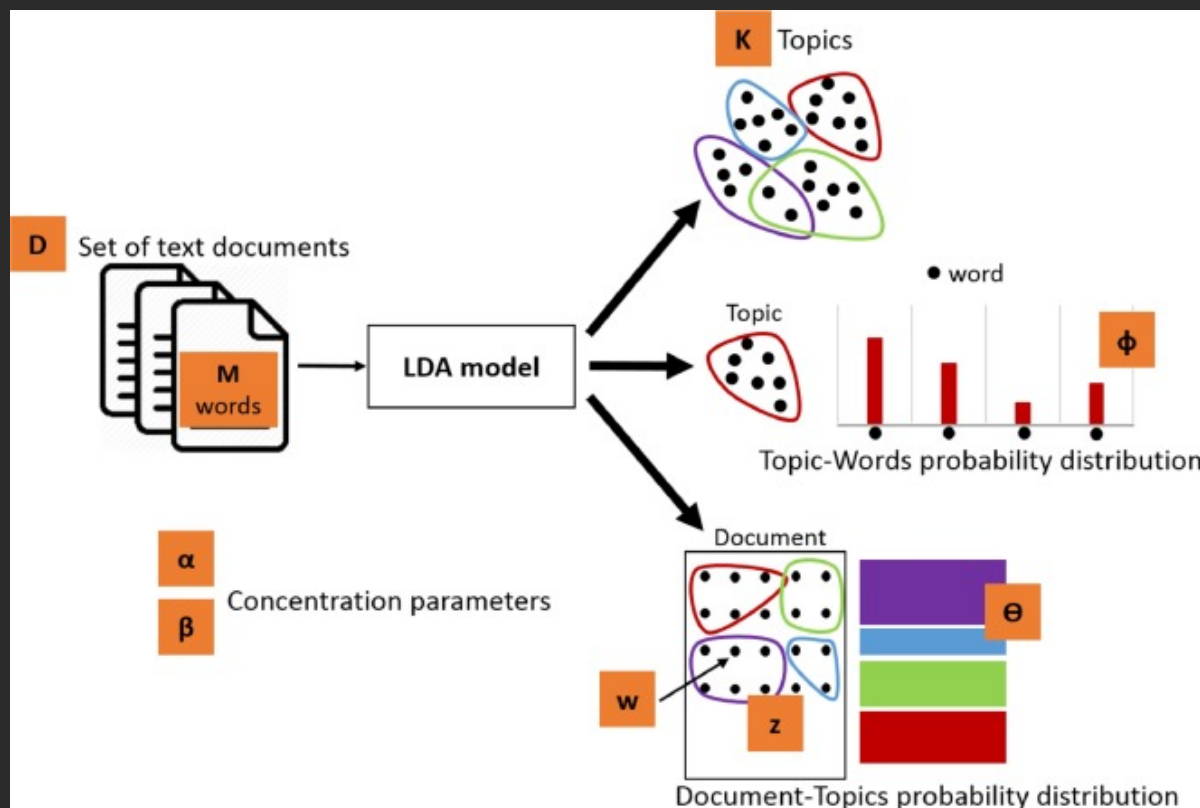


MACHINE LEARNING

- Machine Learning
 - *Top 400 words were seleted and 140 words were mannually removed (e.g. he, did, go, yo, yeserday, etc). Final key words: 260*
 - Latent Dirichlet Allocation (LDA): Unsupervised learning for clusterings/topics
 - Grid search: $n = 2$ clusterings
 - From 2 up to 8 clusterings: human choice of 5 clusterings



LATENT DIRICHLET ALLOCATION MODEL EXPLAINED



LATENT DIRICHLET ALLOCATION MODEL EXPLAINED

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
       2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```



LATENT DIRICHLET ALLOCATION MODEL EXPLAINED

	Topic0	Topic1	Topic2	Topic3	Topic4	dominant_topic
Tweet_0	0.17	0.02	0.02	0.34	0.44	4
Tweet_1	0.03	0.86	0.03	0.03	0.03	1
Tweet_2	0.04	0.04	0.04	0.84	0.04	3
Tweet_3	0.02	0.02	0.02	0.91	0.02	3
Tweet_4	0.04	0.04	0.04	0.04	0.84	4



MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

1. School_reopening (n=8557): school play football back week season colleg kid player sport cancel fall team game health risk plan check children high concern big athlet year take wonder free open reopen confer

MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

1. School_reopening (n=8557): school play football back week season colleg kid player sport cancel fall team game health risk plan check children high concern big athlet year take wonder free open reopen confer

2. 2020_Election (n=7945): trump realdonaldtrump american vote respons peopl presid elect america kill countri mail help state million think unemploy relief world money economi order call democrat gop take biden lie fund hous

MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

1. *School_reopening (n=8557)*: school play football back week season colleg kid player sport cancel fall team game health risk plan check children high concern big athlet year take wonder free open reopen confer

2. *2020_Election (n=7945)*: trump realdonaldtrump american vote respons peopl presid elect america kill countri mail help state million think unemploy relief world money economi order call democrat gop take biden lie fund hous

3. *State_News_Economy (n=7184)*: test new time case death posit state thank today well number report job work amp result rate first back florida hospit quarantin neg

MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

1. *School_reopening (n=8557)*: school play football back week season colleg kid player sport cancel fall team game health risk plan check children high concern big athlet year take wonder free open reopen confer

2. *2020_Election (n=7945)*: trump realdonaldtrump american vote respons peopl presid elect america kill countri mail help state million think unemploy relief world money economi order call democrat gop take biden lie fund hous

3. *State_News_Economy (n=7184)*: test new time case death posit state thank today well number report job work amp result rate first back florida hospit quarantin neg

4. *Policies_Mask_Safty (n=8650)*: amp like mask peopl wear feel social student commun work spread safe help real everyon post love distanc put great famili face

MACHINE LEARNING CONTINUED

- Human Choice of 5 Clusterings:

1. *School_reopening (n=8557)*: school play football back week season colleg kid player sport cancel fall team game health risk plan check children high concern big athlet year take wonder free open reopen confer

2. *2020_Election (n=7945)*: trump realdonaldtrump american vote respons peopl presid elect america kill countri mail help state million think unemploy relief world money economi order call democrat gop take biden lie fund hous

3. *State_News_Economy (n=7184)*: test new time case death posit state thank today well number report job work amp result rate first back florida hospit quarantin neg

4. *Policies_Mask_Safty (n=8650)*: amp like mask peopl wear feel social student commun work spread safe help real everyon post love distanc put great famili face

5. *General_Critics (n=7664)* die peopl month year vaccin fuck home live shit good take pleas sick work life mani busi flu last first right stop kid think patient care immun

FUTURE DIRECTIONS

- COVID-19 and cancer patients
- Cancer and financial hardship
- Historical data
- Model improvements
- Looking for future updates? Click on: [Twitter Sentiment](#)

HELP US FIGHT CANCER

If you find this analysis helpful, please click on our logo and make a donation:



QUESTIONS AND COMMENTS

RESULTS

Rank 2020

act

✓ Click to Show All
Values



merican

1 amp

35 ass

athlet

back

bad

believ

best

better

biden

big

bill

black

blame

break

